2Г

УДК 681.3: 519.68

А.Я. Гладун[1], Ю.В. Рогушина[2]
[1]Міжнародний науково-учбовий центр інформаційних технологій і систем НАНУ та МОНУ, м. Київ, Україна
glanat@yahoo.com
[2]Інститут програмних систем НАНУ, м. Київ, Україна
jjj@ukr.net

Основи методології формування тезаурусів з використанням онтологічного та мереологічного аналізу

У роботі розглянуто актуальні напрямки подання розподілених знань, які базуються на онтологічному аналізі, використання онтологій та тезаурусів в сучасних Інтернет-технологіях. Викладено теоретичні положення, які використовують елементи мереологічного та онтологічного аналізу, та їх практичне застосування у створенні визначень термінів тезауруса ПрО.

Вступ

Сьогодні головні напрямки розвитку інформаційних технологій (ІТ) пов'язані зі створенням інформаційних систем, базованих на знаннях у відповідних предметних областях (ПрО). Значну кількість людей можна вважати експертами у певних ПрО, які відображають, наприклад, їхню професійну чи наукову діяльність або інші інтереси. Діяльність окремих людей і організацій все більше залежить від наявності у них *знань* – одного із найцінніших ресурсів сучасного суспільства – і здатності ефективно їх використовувати. Однак існуючі засоби представлення знань ще є досить недосконалими і часто змушують людей знову і знову шукати вирішення одних і тих самих задач [1]. У цій роботі розглянуті актуальні напрямки подання розподілених знань, які базуються на онтологічному аналізі, використання онтологій та тезаурусів в сучасних Інтернет-технологіях, теоретичні положення, які використовують елементи мереологічного та онтологічного аналізу, та їх практичне застосування у створенні визначень термінів тезауруса ПрО.

Постановка задачі

Тезауруси та онтології – інтероперабельні засоби подання знань ПрО, які придатні для використання розподіленими Web-застосуваннями, – потребують методів та засобів їх адекватного створення. **Мета роботи** полягає у тому, щоб коректно визначити відношення між різними термінами тезауруса, для чого пропонується використовувати основи мереологічного підходу. Крім того, пропонуються правила створення визначень термінів тезауруса, які дозволяють формувати більш якісні знання про ПрО, тобто більш строго проаналізувати відношення, що використовуються в онтологічному аналізі, зокрема, мереологічні відношення.

Онтологічний підхід до подання знань ПрО

Важливо досягти інтероперабельності знань, які використовуються в інформаційних системах (ІС), тобто знання, які отримані при створенні однієї ІС, повинні бути придатними для роботи інших ІС. Онтології є саме такою формою подання знань. Наприклад, у проекті Semantic Web онтології є однією з базових технологій, спрямованих на перетворення Web у глобальну розподілену базу знань.

Онтологія – угода про спільне використання понять (термінів), яка містить засоби подання предметних знань і домовленості про методи логічного виведення. Вона може розглядатися як формалізований опис погляду на світ у конкретній сфері інтересів, який складається з набору термінів і правил використання цих термінів, які обмежують їхнє значення в рамках конкретної ПрО [2].

Формальна модель онтології ПрО – це упорядкована трійка O=<T, R, F>, де T – скінчена множина термінів ПрО, які описує онтологія O; R – скінчена множина відносин між термінами заданої ПрО; F – скінчена множина функцій інтерпретації, заданих на термінах і/чи відносинах онтології O. Якщо множина термінів онтології цілком визначається її ПрО, то множина використовуваних в онтологіях відношень більш незалежна від ПрО. Відношення представляють тип взаємодії між концептами домену. При створенні онтологій найбільшу складність представляє формування множини F, тому що цей процес вимагає застосування спеціальних навичок з області інженерії знань і формальної логіки [3]. Основна робота з формування онтологій припадає на формування множини X, причому ця робота доступна більшості фахівців довільної предметної області. Трохи складніше визначити множину відносин R, яку треба використовувати для моделювання ПрО, але в більшості випадків можна використовувати стандартні набори з 10 – 20 базових відносин («бути частиною», «бути підкласом», «бути однаковим» тощо).

У [4] виділені найбільш загальні відношення в реальних доменах, а саме: зв’язок еквівалентності, таксономічний зв’язок, структурний зв’язок, зв’язок залежності, топологічний зв’язок, зв’язок причини і наслідку, функціональний зв’язок, хронологічний зв’язок, зв’язок подоби, умовний зв’язок і цільовий зв’язок. Однак не всі ці онтологічні відносини рівноцінні, у цьому наборі відносин можна виділити трохи найбільш фундаментальних, таких як: таксономія і мереологія.

Відношення називають фундаментальним, якщо на базі одного даного відношення може бути побудована формальна система, яка дозволяє виражати основні математичні поняття. Існує чотири фундаментальних відношення, на базі кожного з яких може бути побудована математика: відношення приналежності (теорія множин ZF і NF), відношення між функцією, її аргументом і результатом (теорія множин фон Неймана), відношення іменування (онтологія Лесьневського) і відношення «частина-ціле» (мереологія) [5].

Таксономія – наука, яка вивчає розподіл на упорядковані групи чи категорії. Типовим прикладом таксономії є ієрархічний список. Класи в онтології звичайно організовуються в таксономії. Існує кілька загальних таксономічних властивостей: 1) асиметрія; 2) транзитивність; 3) антирефлексивність.

Якщо термін «таксономія» досить широко розповсюджений і не вимагає додаткового визначення, то термін «мереологія» у дослідженнях, пов’язаних з інформаційними технологіями, застосовується значно рідше і має потребу в додатковому поясненні. **Мереологія** (від грецького «частина» і «вивчення») – формальна теорія про частини і зв’язані з ними поняття. Цей термін використовувався польським

філософом С. Лесьневським [6], який аналізує філософські, логічні і математичні складові основ математики. Відношення «*частина-ціле*» є винятково важливим тому, що воно утворює основу поняття *системи*, яка часто використовується в сучасному науковому пізнанні. У мереології відношення «є частиною» (*part-of*) має наступні властивості: 1. Асиметрія: $(x << y) \Rightarrow \neg(y << x)$; 2. Транзитивність: $(x << y) \wedge (y << z) \Rightarrow x << z$. Ми використовуємо наступні аксіоми мереології.

Аксіома 1. Існує пуста система $\exists s \forall x, y, c \neg M(x, y, c) = \varnothing$, причому така система є єдиною.

Аксіома 2. Система, яка має хоча б одну частину, сама є власною частиною: $\forall s(\exists x, y, cM(x, y, c, s) \Rightarrow \exists zM(s, s, z, s))$.

Аксіома 3. Якщо інших зв'язків у двох об'єктів може бути декілька, то наявність пустого зв'язку заперечує можливість будь-якого іншого зв'язку:

$$\forall x, y, s(M(x, y, \varnothing, s) \Rightarrow \forall c(M(x, y, c, s \vee M(y, x, c, s) \Rightarrow c = \varnothing).$$

Мереологія виходить за межі вивчення часткових відносин між елементами загальних систем. Вона також займається тими об'єктами, частини яких релевантні цілому. Такі об'єкти ідентифікуються як екземпляри.

Для мереології виконується *аксіома об'ємності:* системи, які зв'язують ті самі об'єкти тими самими способами, рівні. Аксіоми мереології описують взаємозв'язки між системою і її елементами. Наприклад, система, яка має хоча б одну частину, сама є власною частиною. Серед мереологічних відношень можна виділити сім різних класів:

1. **Компонент-Об'єкт:** сторінка-книга.
2. **Член-Колекція**: дерево-ліс.
3. **Частина-Маса:** шматок-хліб.
4. **Матеріал-Об'єкт:** алюміній-літак.
5. **Властивість-Діяльність:** бачити-читати.
6. **Стадія-Процес:** заварювання-готування сподіваючись.
7. **Місцевість-Область:** Закарпаття-Україна.

Більшість досліджень відношень «частина-ціле» присвячені вивченню частин, але можна ідентифікувати різні типи цілого відповідно до наступних властивостей:

1. Чи віддільна частина від цілого («мелодія-пісня» чи «вагон-потяг»).
2. Є частини просторовими чи часовими («кімната-квартира» чи «зима-рік»);
3. Чи грає частина певну функціональну роль відносно цілого («двигун-автомобіль»).
4. Чи є частини неподільними («атом-молекула»).

Створюючи онтології, необхідно чітко визначати, до якого типу відношень належать ті відношення, які встановлюються між термінами. Помилкова класифікація може негативно вплинути на якість онтологій.

Онтології та тезауруси

Для вирішення багатьох застосовних задач в якості бази знань можна замість онтологій використовувати тезауруси. Донедавна терміни «онтологія» і «тезаурус» використовувалися як синоніми, однак зараз у ІТ тезаурус частіше застосовують для опису лексики в проекції на семантику, а онтологію – для моделювання семантики і прагматики в проекції на мову подання [7]. Звичайно тезаурус Т визначають як словник, який містить лексичні одиниці (ЛО) із вказівкою семантичних зв'язків між ними.

Слово «*тезаурус»* походить від грецького «*скарбниця»*. Вперше його застосував у значенні, близькому до сьогоднішнього, ще в XIII столітті Б. Датіні в енциклопедії «Книга про скарб». Згідно з «Сучасним словником іноземних слів»: *тезаурус* – 1) словник, у якому максимально повно подані усі слова мови з вичерпним переліком прикладів їхнього вживання в текстах; у повному обсязі це може бути реалізовано лише для мертвих мов; 2) ідеографічний словник, у якому показані семантичні відносини (синонімічні, родо-видові тощо) між лексичними одиницями; 3) в інформатиці – повний систематизований набір даних про будь-яку галузь знань, яка дозволяє людині чи комп'ютеру в ній орієнтуватися. Тезаурус (відповідно до третього визначення) можна розглядати як окремий випадок онтології, який значно простіше формувати та аналізувати, але він, звичайно, не дозволяє подавати певні більш складні структури знань [8].

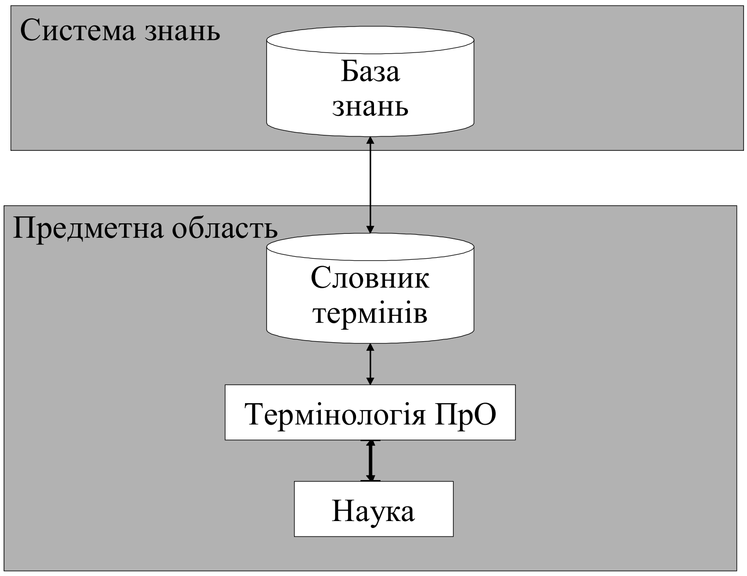


Рисунок 1 – Зв'язок предметної області й системи знань

Багато авторів розглядають тезаурус як ієрархічну структуру або ряд ієрархічних структур, у які групуються групи понять із загальними характеристиками, які виражені термінами природної мови. Для більшості реальних ПрО це не так, оскільки різні відношення, використовувані в ПрО, непорівнювані (наприклад, «стаття написана автором», «програма розроблена мовою»). Більш коректно тезаурус можна представити як семантичну мережу. У вузлах такої мережі перебувають терміни, зв'язані відносинами з обмеженого набору. Для подання тезауруса доцільно також використати орієнтовані графи й фрейми. Наслідуючи багато достоїнств семантичної мережі, фреймова мережа дозволяє у якості вершин представити складні структури (фрейми), які мають, зокрема, незаповнені поля (слоти), що дає нові можливості при описі вкладених структур, перемиканні між різними застосуваннями тощо.

Формальна модель тезауруса $Ts = \langle T, R \rangle$, де T – скінчена множина термінів, а R – скінчена множина відношень між цими термінами. **Термін** – це слово або словесний комплекс, який співвідноситься з поняттям визначеної організованої області пізнання (науки, техніки) та вступає в системні відносини з іншими словами і словесними комплексами і утворює разом з ними в будь-якому окремому випадку чи у певний

час замкнуту систему, яка відрізняється високою інформативністю, однозначністю, точністю й експресивною нейтральністю. Множина термінів тезауруса T відповідає множині концептів X онтології O.

Виділяють тезауруси інформаційних ресурсів (ІР), створені в результаті обробки термінології певного інформаційного ресурсу та встановлення зв'язків між відповідними термінами; тезауруси ПрО, які об'єднують тезауруси відповідних ІР та знання експертів ПрО; тезауруси особистості (користувача, експерта тощо), які містять терміни, знайомі людині внаслідок обробки відповідних ІР або з інших джерел (рис. 2).

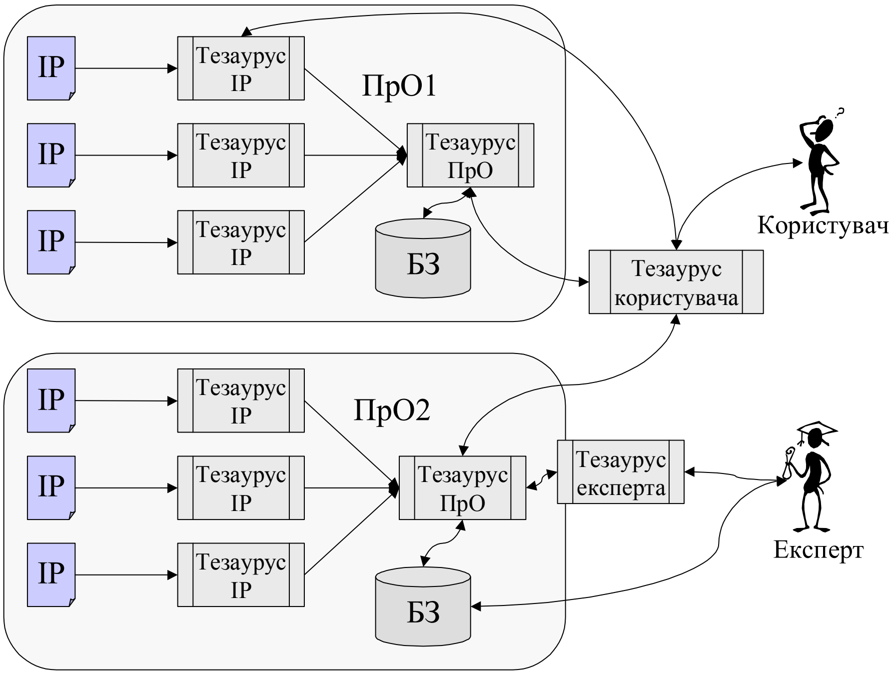


Рисунок 2 – Типи тезаурусів

Основні технологічні фази створення тезауруса: 1. Виділення ЛО – формування словника T; 2. Розробка набору семантичних зв'язків; 3. Актуалізація зв'язків – установлення зв'язків між термінами. При актуалізації семантичних зв'язків між термінами тезауруса можна використовувати знання експертів, а також документи, призначені як для фіксації структури знань ПрО (словники, класифікатори тощо), так і самі знання, які відображають ПрО (реферати, статті, монографії тощо). Згідно з методологією розробки онтологічних моделей IDEF5 [9], [10], побудова тезауруса складається з п'яти основних дій.

1. *Вивчення і систематизація початкових умов* – мети і контексту розробки тезауруса, визначення границь ПрО, яка цікавить користувача.

2. *Збір і накопичення даних* – підбір ІР, які відносяться до цієї ПрО.

3. *Аналіз даних* – вивчення відібраних ІР, формування словника термінів ПрО, які містяться у відібраних ІР.

4. *Початкова розробка тезауруса* – встановлення зв'язків між термінами ПрО.

5. *Уточнення і затвердження тезауруса* – аналіз користувачем отриманого тезауруса та його коректування.

Крім цих загальних методичних вказівок, доцільно дотримуватися низки практичних рекомендацій, спрямованих на створення якісного тезауруса.

Практичні рекомендації стосовно визначення термінів тезауруса

Визначення термінів тезауруса можуть формуватися автоматично (шляхом аналізу повнотекстових документів та інших інформаційних джерел), витягатися з інших баз знань (тезаурусів, онтологій тощо) або створюватися вручну експертом ПрО [11-14]. У всіх випадках доцільно дотримуватися певних правил та вдосконалювати отримані визначення відповідно до наведених нижче принципів. **Поняття** – одиниця інформації, створена як унікальна комбінація характеристик. **Визначення** – представлення поняття через описовий вираз, що дозволяє відрізнити його від інших пов'язаних з ним понять.

Визначення терміна повинне:

1) бути викладене в однині: виключення складають поняття, які самі є множинними. Приклад: «Ідентифікаційний код»: 1) добре визначення «Число, що ідентифікує людину»; 2) погане визначення «Число для ідентифікації людей». Причина – у поганому визначенні використовується слово «люди», яке є неоднозначним, тому що це можна зрозуміти так, ніби один номер може посилатися на кілька людей;

2) визначати, чим є наведене поняття, а не тільки чим воно не є: приклад: «Розмір вартості доставки»: 1) добре визначення «Розмір витрат, які несе вантажовідправник для переміщення товарів з одного місця до іншого»; 2) погане визначення «Розмір витрат, які не відносяться до витрат на пакування, оформлення, завантаження, розвантаження та страхування». У поганому прикладі не вказано, що входить до поняття елементу даних;

3) мати вигляд описової фрази або речення: речення (у більшості мов) необхідне для формування точного визначення, яке містить важливі характеристики поняття. Просте наведення одного або кількох синонімів не є достатнім. Якщо потрібно більше, ніж описова фраза, слід використовувати повні, граматично правильні речення. Приклад – «Ім'я агента»: 1) добре визначення «Назва сторони, яка уповноважена діяти від імені іншої сторони»; 2) погане визначення «Представник» Причина – «Представник» є близьким синонімом імені елементу даних, який не може бути адекватним визначенням;

4) містити лише широко відомі скорочення: Розуміння значення скорочення, зокрема абревіатур та ініціалів, зазвичай обмежується певним середовищем. В іншому середовищі ті ж самі скорочення можуть викликати невірне розуміння або непорозуміння. Таким чином, для запобігання неоднозначності, у визначеннях використовуються тільки повні слова без скорочень. Виняток з цієї вимоги може бути зроблений тоді, якщо скорочення є загальновживаним, наприклад, «т.д.» або якщо скорочення є більш зрозумілим, ніж повна форма складного терміна та використовується як окремий термін, наприклад, «радар» (radar – radio detecting and ranging). Усі абревіатури мають бути розшифровані під час їх першої появи.

Приклад 1: «Висота над рівнем моря»: 1) добре визначення: «Вертикальна відстань від середнього рівня моря (СРМ) до рівня, що описується»; 2) погане визначення «Вертикальна відстань від СРМ до рівня, що описується». Погане визначення є незрозумілим, тому що абревіатура СРМ не є загальнозрозумілою і деякі користувачі змушені будуть звертатися до інших джерел для з'ясування її значення. Без зазначення повного слова пошук терміна у словнику може бути складним або взагалі неможливим.

Приклад 2: «Прилад для вимірювання щільності»: 1) добре визначення «Прилад, який використовується для вимірювання концентрації рідини, в одиницях виміру маси до одиниці об'єму (м.д.о.) (тобто фунтів на кубічний фут; кілограмів на кубічний метр)»; 2) погане визначення «Прилад, який використовується для вимірювання концентрації рідини в термінах м.д.о. (тобто фунтів на кубічний фут; кілограмів на кубічний метр)». Причина – м.д.о не є загальновідомим скороченням і його значення може бути незрозумілим для деяких користувачів. Скорочення має бути представлене повними словами;

5) бути викладеним без використання визначень інших даних або базових понять: визначення другого елементу даних або пов'язаного з ним поняття не може з'являтися у визначенні первинного елементу даних. Визначення термінів має наводитись у відповідному глосарії. Якщо потрібне друге визначення, воно має додаватись як примітка після тексту первинного визначення або як окремий запис у словнику. Пов'язані визначення можна отримати за допомогою атрибутів посилання (тобто перехресних посилань).

Приклад 1: «Код типу зразка»: 1) добре визначення «Код, який ідентифікує тип зразка»; 2) погане визначення «Код, який ідентифікує тип обраного зразка. Зразок – це мала частка, вилучена для проведення експериментів. Він може бути як єдиним зразком для тестування, так і сурогатним зразком для контролю якості. Зразок для контролю якості – це сурогатний зразок, обраний для перевірки результатів тестування єдиних зразків». Погане визначення містить два додаткових визначення. Це є визначення «зразка» та «зразка для контролю якості».

Приклад 2: «Емісійний номер банківського кредиту»: 1) добре визначення «Номер посилання, який призначається емісуючим банком документарному кредиту»; 2) погане визначення «Номер посилання, який призначається під час емісії банком документарного кредиту. Документарний кредит – це документ, в якому банк зазначає, що ним був виданий документарний кредит, за яким позичальник отримає платіж, акцепт або залік відповідно до визначених умов та за умови представлення визначених документів». У поганому визначенні міститься визначення, яке повинне міститися у глосарії;

6) відображати суттєвий зміст поняття: усі первинні характеристики поняття, що представляється, мають бути відображені у визначенні з відповідним рівнем специфічності залежно від контексту. При цьому необхідно запобігати зазначення неважливих параметрів. Потрібний рівень деталізації залежить від потреб користувача системи та середовища.

Приклад 1: «Номер послідовності завантаження вантажу» (визначений контекст: будь-яка форма транспортування):

1) добре визначення: Номер, який вказує послідовність, в якій здійснюється завантаження до транспортного засобу або елемента транспортного середовища;

2) погане визначення: Номер, який відображає послідовність, в якій здійснюється завантаження до вантажівки.

У даному контексті вантажі можуть транспортуватись різними транспортними засобами, приміром, вантажівками, судами, вантажними потягами. Транспортування не обмежене лише вантажівками.

Приклад 2: «Сума за рахунком-фактурою»:

1) добре визначення: Загальна сума, яку потрібно сплатити за рахунком-фактурою;

2) погане визначення: Загальна сума вартості всіх елементів, зазначених в рахунку-фактурі, включаючи усі відрахування, зокрема знижки та дисконти, та додаткові платежі, зокрема страхові, транспортні та накладні витрати тощо. У поганому визначенні міститься зайва інформація;

7) бути точним та однозначним: точний зміст та інтерпретація поняття, яке визначається, мають бути зрозумілі з визначення. Визначення має бути достатньо ясним, щоб забезпечити його однозначну інтерпретацію.

Приклад: «Дата Отримання Вантажу»: 1) добре визначення «Дата, на яку вантаж передається отримувачу»; 2) погане визначення «Дата, на яку здійснюється доставка вантажу». У поганому визначенні не роз'яснюється, що таке «доставка». Під «доставкою» можна зрозуміти як момент розвантаження товару у певному місті, так і факт передачі товару кінцевому отримувачу. Не виключено, що кінцевий отримувач ніколи не отримає вантаж, або його передача може здійснитися через кілька днів після розвантаження;

8) бути коротким: слід запобігати використання додаткових фраз описового характеру. Приклад: «Ім'я набору символів»: 1) добре визначення «Ім'я, що присвоюється набору фонетичних або ідеографічних символів, у які зашифровані дані»; 2) погане визначення «Ім'я, що присвоюється набору фонетичних або ідеографічних символів, у яких зашифровані дані для забезпечення використання цього реєстру метаданих або, якщо говорити про загальний вжиток, спроможність системного обладнання і програмного забезпечення обробляти дані, зашифровані одним або декількома шифрами». У поганому визначенні усі фрази після виразу «у яких зашифровані дані» є зайвими описовими фразами;

9) мати можливість використовуватися окремо: зміст поняття має бути наочним у визначенні. Для розуміння поняття не повинні бути потрібні додаткові роз'яснення.

Приклад: «Назва Міста Розміщення Школи»:

1) добре визначення: Назва міста, де знаходиться школа;

2) погане визначення: Див. сайт школи.

Причина – погане визначення не є самостійним, необхідно звернутися до додаткового джерела (сайта школи) для розуміння поняття;

10) бути поданим без використання пояснювальної інформації, функціонального використання або процедурної інформації: у разі потреби такі пояснення можуть бути розміщені у інших атрибутах метаданих;

11) запобігати циклічних посилань: два визначення не повинні бути визначені одне через одне.

Приклад: два елементи даних з поганими визначеннями: 1) «Ідентифікаційний Номер працівника – Номер, що призначається працівнику»; 2) «Працівник – Людина, яка має відповідний ідентифікаційний номер працівника». Визначення посилаються одне на одне. При цьому в жодному визначенні не наведено зміст поняття;

12) використовувати ту ж саму термінологію та логічну структуру для пов'язаних визначень: для близьких або пов'язаних визначень має використовуватись одна й та ж сама термінологія та синтаксис.

Приклад: Даний приклад ілюструє цю ідею. Обидва визначення відносяться до пов'язаних понять і тому мають однакову логічну структуру і близьку термінологію. 1) «Дата відправлення товарів» – Дата, у яку товари були відправлені даній стороні; 2) «Дата отримання товарів» – Дата, у яку товари були отримані даною стороною. Використання єдиної термінології та синтаксису значно спрощує розуміння. Інакше користувачі можуть не зрозуміти, чому для пов'язаних визначень використовуються різні терміни.

Висновки

Запропоновані у статті методологічні рекомендації доцільно використовувати для створення нових та вдосконалення існуючих тезаурусів різних предметних областей. У перспективі слід розробити та прийняти державні та міжнародні стандарти, пов'язані з правилами надання визначень термінів та створення тезаурусів і онтологій.

Література

1. Рогушина Ю.В., Гладун А.Я Використання онтології як засобу інтеграції знань про інформаційну систему // Відбір і обробка інформації. – 2006, Львів: ФМІ ім. Г.В. Карпенка. – № 5, Вип. 24 (100). – С. 43-49.
2. Гладун А.Я., Рогушина Ю.В. Онтологический подход к поиску веб-сервисов в распределенной среде Интернет // Информатика. – Минск, 2006. – № 4. – С. 116-127.
3. Knowledge Maps: An essential technique for conceptualisation / A. Gómez-Pérez, A. Moreno, J. Pazos, A. Sierra-Alonso // Data & Knowledge Engineering. – 2000. – V. 33(2). – P. 169-190.
4. Непейвода Н.Н. Мереология. – Режим доступа: http://www.logic.ru/Russian/events/ifras/nnn.pdf
5. Лесьневский С. Об основаниях математики. – Режим доступа: http://www.philosophy.ru/library/logic/lesnewxi.html.
6. Рогушина Ю.В., Гладун А.Я. Мереологические аспекты онтологического анализа интеллектуальных Web-сервисов // Збірник праць VII Міжнародної конференції «Інтелектуальний аналіз інформації» IAI-2007 Київ. – 15-18 травня 2007 р. – С. 312-321.
7. Нариньяни А.С. Кентавр по имени ТЕОН: Тезаурус + Онтология. – Режим доступа: http://www.artint.ru/articles/narin/teon.htm.
8. Гладун А.Я., Рогушина Ю.В. Формирование тезауруса предметной области, как средства моделирования информационных потребностей пользователя при поиске в Интернет // Вестник компьютерных и информационных технологий. – Москва: Машиностроение, 2007. – № 1. – С. 56-68.
9. Верников Г. Стандарт онтологического исследования IDEF5. – Режим доступа: http://www.vernikov.ru/material36.html.
10. Браславский П.И., Гольдштейн С.Л., Ткаченко Т.Я. Тезаурус как средство описания систем знаний // Информационные процессы и системы. – 1997. – № 11, Серия 2. – С. 16-22.
11. Андреев А.М., Березкин Д.В., Симаков К.В. Особенности проектирования модели и онтологии предметной области для поиска противоречий в правовых электронных библиотеках. – Режим доступа: http://www.inteltec.ru/publish/articles/textan/RCDL2004.shtml
12. Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. – Спб.: Питер, 2001.
13. Гладун А.Я., Рогушина Ю.В. Онтологии и мультилингвистические тезаурусы как основа семантического поиска информационных ресурсов Интернет // The Proc. of XII-th Intern. Conf. KDS'2006. – Varna, Bulgaria. – P. 115-121.
14. Noy N., Musen M. The PROMPT Suite: Interactive Tools For Ontology Merging And Mapping // Stanford Medical Informatics. – Stanford Univ., 2003.
15. Gruber T.R. A translation approach to portable ontology specifications // Knowledge Acquisition. – 1993. – V. 5. – P. 199-220.

А.Я. Гладун, Ю.В. Рогушина

Основы методологии формирования тезаурусов с использованием онтологического и мереологического анализа

В этой работе рассмотрены актуальные направления представления распределенных знаний, которые базируются на онтологическом анализе, использовании онтологий и тезаурусов в современных Интернет-технологиях. Изложены теоретические положения, которые используют элементы мереологического и онтологического анализа, а также их практическое применение при создании определений терминов тезауруса ПрО.

A. Gladun, J. Rogushina

Bases of Methodology of Formation of Thesauruses with Use Ontologic and the Mereologic Analysis

In this paper actual directions of representation of the distributed knowledge which are based on the ontologic analysis, use ontologies and thesauruses in modern Internet-technologies are considered. Theoretical positions which use elements mereologic and the ontologic analysis, and also their practical application at creation of definitions of terms of the thesaurus domain are stated.

Стаття надійшла до редакції 17.07.2008.